Exhibit I

NIH Organization

Begun as a one-room Laboratory of Hygiene in 1887, the National Institutes of Health (NIH) today is one of the world's foremost medical research centers. An agency of the Department of Health and Human Services, the NIH is the Federal focal point for health research. Read more »

NIH Office of the Director

The Office of the Director (OD) is responsible for setting policy for NIH and for planning, managing, and coordinating the programs and activities of all 27 of NIH's Institutes and Centers. The OD program offices include the Office of AIDS Research, Office of Behavioral and Social Sciences Research, Office of Disease Prevention, The Office of Strategic Coordination and Office of Research on Women's Health, among others.

NIH Institutes

- National Cancer Institute (NCI)
- National Eye Institute (NEI)
- National Heart, Lung, and Blood Institute (NHLBI)
- National Human Genome Research Institute (NHGRI)
- National Institute on Aging (NIA)
- National Institute on Alcohol Abuse and Alcoholism (NIAAA)
- National Institute of Allergy and Infectious Diseases (NIAID)
- National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS)
- National Institute of Biomedical Imaging and Bioengineering (NIBIB)
- *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD)
- National Institute on Deafness and Other Communication Disorders (NIDCD)
- National Institute of Dental and Craniofacial Research (NIDCR)
- National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)
- National Institute on Drug Abuse (NIDA)
- National Institute of Environmental Health Sciences (NIEHS)

- National Institute of General Medical Sciences (NIGMS)
- National Institute of Mental Health (NIMH)
- National Institute on Minority Health and Health Disparities (NIMHD)
- National Institute of Neurological Disorders and Stroke (NINDS)
- National Institute of Nursing Research (NINR)
- National Library of Medicine (NLM)

NIH Centers

- Center for Information Technology (CIT)
- Center for Scientific Review (CSR)
- Fogarty International Center (FIC)
- National Center for Advancing Translational Sciences (NCATS)
- National Center for Complementary and Integrative Health (NCCIH)
- NIH Clinical Center (CC)

This page last reviewed on June 14, 2018

Search the Almanac

Exhibit J

Office of the Director, NIH

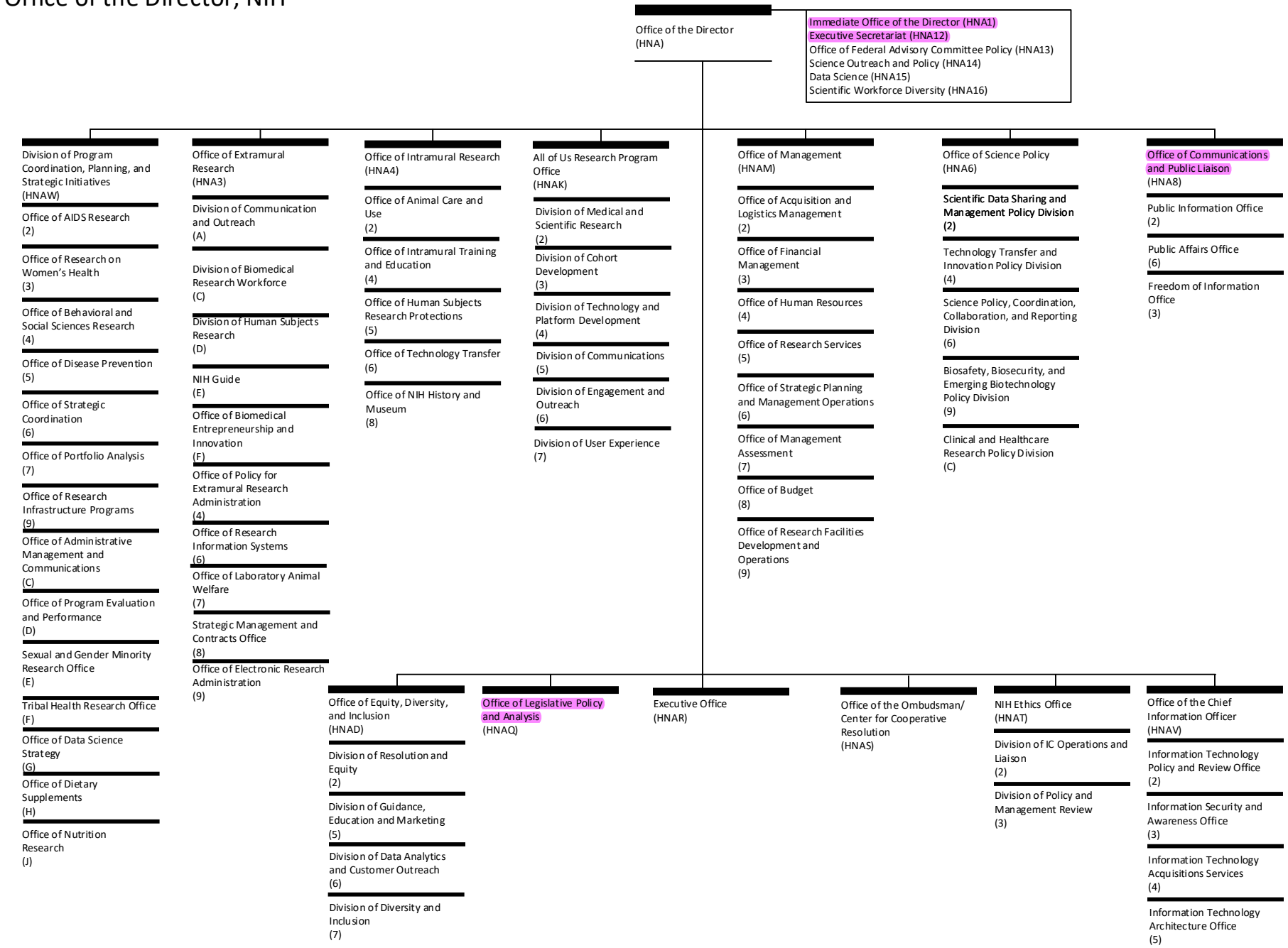


Exhibit K

Office of Communications and Public Liaison

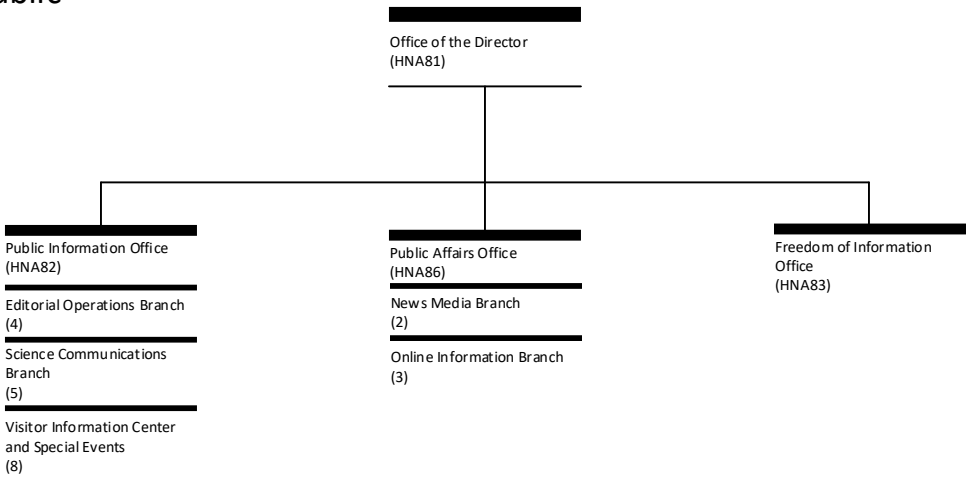
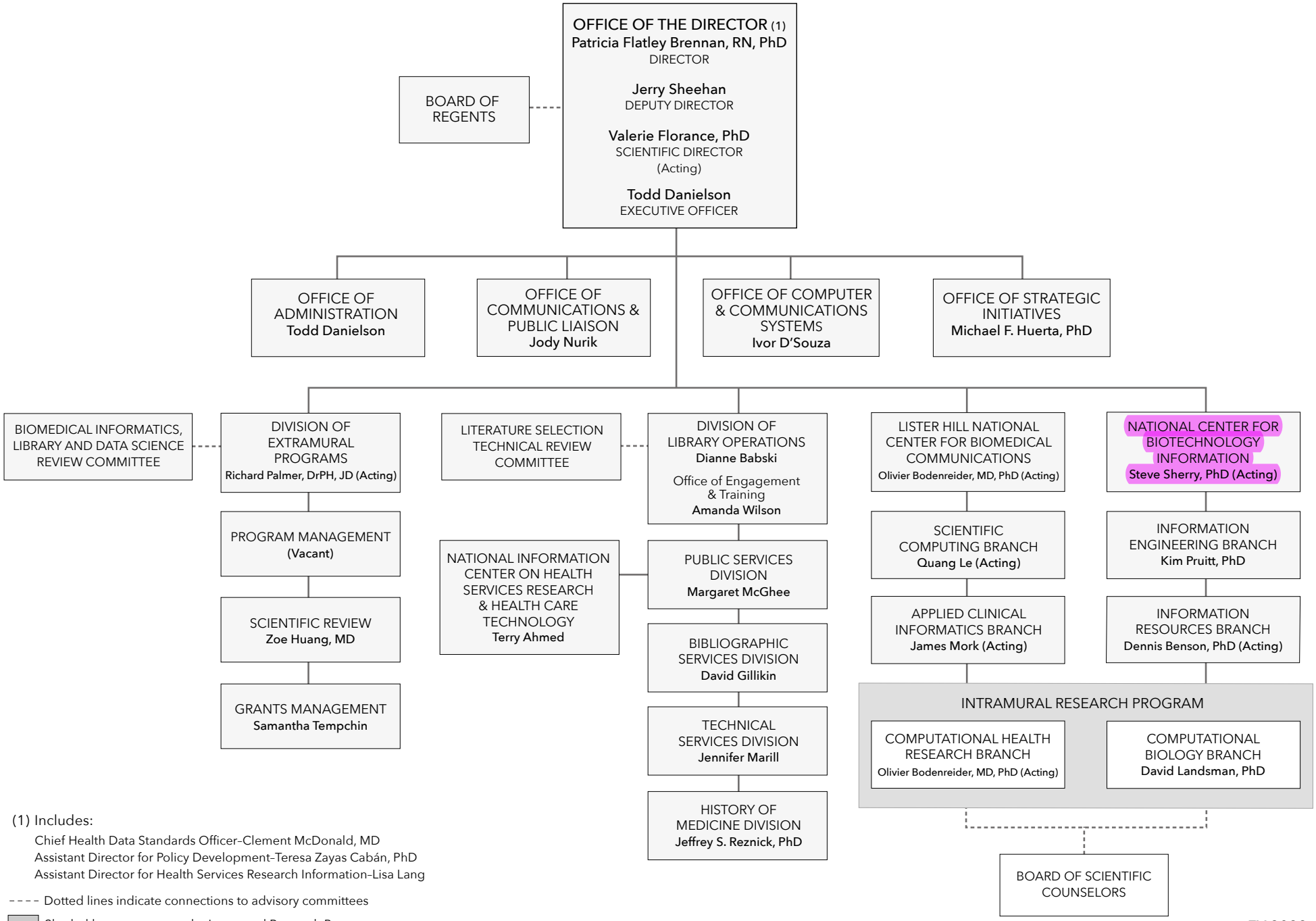


Exhibit L

NATIONAL LIBRARY OF MEDICINE



(1) Includes:

Chief Health Data Standards Officer–Clement McDonald, MD
Assistant Director for Policy Development–Teresa Zayas Cabán, PhD
Assistant Director for Health Services Research Information–Lisa Lang

---- Dotted lines indicate connections to advisory committees

Shaded boxes represent the Intramural Research Program

FY 2022

Last Reviewed June 2022

Exhibit M

NLM By Organization

Also see: NLM Organization Chart (PDF) or text version

- **Office of the Director** (OD)
  (Advisory Body: **NLM Board of Regents**)
  - Office of Administration
  - Office of Communications and Public Liaison
  - Office of Computer & Communications Systems (OCCS)
    - Applications Branch
    - Medical Language Branch
    - Systems Technology Branch
  - Office of Strategic Initiatives
  - **Library Divisions**
- **Division of Extramural Programs**
  (Advisory Body: **Biomedical Library and Informatics Review Committee**)
  - Scientific Review Office
  - Grants Management Office
  - Program Office
- **Division of Library Operations**
  (Advisory Body: **NLM Literature Selection Technical Review Committee**)
  - Bibliographic Services Division
    - Index Section
    - MEDLARS Management Section
  - History of Medicine Division
  - Office of Engagement & Training
  - Public Services Division
    - Collection Access Section
    - Preservation and Collection Management Section
    - Reference and Web Services
      - National Information Center on Health Services Research (NICHSR)
  - Technical Services Division
    - Cataloging and Metadata Management Section
    - Collection Development and Acquisitions Section
    - Library Technology Services Section
- **Lister Hill National Center for Biomedical Communications**
  - Applied Clinical Informatics Branch
  - Computational Health Research Branch
  - Scientific Computing Branch
- **National Center for Biotechnology Information** (NCBI)
  - Computational Biology Branch
  - Information Engineering Branch
  - Information Resources Branch

Last Reviewed: June 24, 2022

Exhibit N

All Resources

All | Databases | Downloads | Submissions | Tools | How To

Databases

Assembly
A database providing information on the structure of assembled genomes, assembly names and other meta-data, statistical reports, and links to genomic sequence data.

BioCollections
A curated set of metadata for culture collections, museums, herbaria and other natural history collections. The records display collection codes, information about the collections' home institutions, and links to relevant data at NCBI.

BioProject (formerly Genome Project)
A collection of genomics, functional genomics, and genetics studies and links to their resulting datasets. This resource describes project scope, material, and objectives and provides a mechanism to retrieve datasets that are often difficult to find due to inconsistent annotation, multiple independent submissions, and the varied nature of diverse data types which are often stored in different databases.

BioSample
The BioSample database contains descriptions of biological source materials used in experimental assays.

Bookshelf
A collection of biomedical books that can be searched directly or from linked data in other NCBI databases. The collection includes biomedical textbooks, other scientific titles, genetic resources such as *GeneReviews*, and NCBI help manuals.

ClinVar
A resource to provide a public, tracked record of reported relationships between human variation and observed health status with supporting evidence. Related information in the NIH Genetic Testing Registry (GTR), MedGen, Gene, OMIM, PubMed and other sources is accessible through hyperlinks on the records.

ClinicalTrials.gov
A registry and results database of publicly- and privately-supported clinical studies of human participants conducted around the world.

Computational Resources from NCBI's Structure Group
A centralized page providing access and links to resources developed by the Structure Group of the NCBI Computational Biology Branch (CBB). These resources cover databases and tools to help in the study of macromolecular structures, conserved domains and protein classification, small molecules and their biological activity, and biological pathways and systems.

Consensus CDS (CCDS)
A collaborative effort to identify a core set of human and mouse protein coding regions that are consistently annotated and of high quality.

Conserved Domain Database (CDD)
A collection of sequence alignments and profiles representing protein domains conserved in molecular evolution. It also includes alignments of the domains to known 3-dimensional protein structures in the MMDB database.

Database of Genomic Structural Variation (dbVar)
The dbVar database has been developed to archive information associated with large scale genomic variation, including large insertions, deletions, translocations and inversions. In addition to archiving variation discovery, dbVar also stores associations of defined variants with phenotype information.

Database of Genotypes and Phenotypes (dbGaP)
An archive and distribution center for the description and results of studies which investigate the interaction of genotype and phenotype. These studies include genome-wide association (GWAS), medical resequencing, molecular diagnostic assays, as well as association between genotype and non-clinical traits.

Database of Short Genetic Variations (dbSNP)

Includes single nucleotide variations, microsatellites, and small-scale insertions and deletions. dbSNP contains population-specific frequency and genotype data, experimental conditions, molecular context, and mapping information for both neutral variations and clinical mutations.

GenBank
The NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis. GenBank consists of several divisions, most of which can be accessed through the Nucleotide database. The exceptions are the EST and GSS divisions, which are accessed through the Nucleotide EST and Nucleotide GSS databases, respectively.

Gene
A searchable database of genes, focusing on genomes that have been completely sequenced and that have an active research community to contribute gene-specific data. Information includes nomenclature, chromosomal localization, gene products and their attributes (e.g., protein interactions), associated markers, phenotypes, interactions, and links to citations, sequences, variation details, maps, expression reports, homologs, protein domain content, and external databases.

Gene Expression Omnibus (GEO) Database
A public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted and tools are provided to help users query and download experiments and curated gene expression profiles.

Gene Expression Omnibus (GEO) Datasets
Stores curated gene expression and molecular abundance DataSets assembled from the Gene Expression Omnibus (GEO) repository. DataSet records contain additional resources, including cluster tools and differential expression queries.

Gene Expression Omnibus (GEO) Profiles
Stores individual gene expression and molecular abundance Profiles assembled from the Gene Expression Omnibus (GEO) repository. Search for specific profiles of interest based on gene annotation or pre-computed profile characteristics.

GeneReviews
A collection of expert-authored, peer-reviewed disease descriptions on the NCBI Bookshelf that apply genetic testing to the diagnosis, management, and genetic counseling of patients and families with specific inherited conditions.

Genes and Disease
Summaries of information for selected genetic disorders with discussions of the underlying mutation(s) and clinical features, as well as links to related databases and organizations.

Genetic Testing Registry (GTR)
A voluntary registry of genetic tests and laboratories, with detailed information about the tests such as what is measured and analytic and clinical validity. GTR also is a nexus for information about genetic conditions and provides context-specific links to a variety of resources, including practice guidelines, published literature, and genetic data/information. The initial scope of GTR includes single gene tests for Mendelian disorders, as well as arrays, panels and pharmacogenetic tests.

Genome
Contains sequence and map data from the whole genomes of over 1000 organisms. The genomes represent both completely sequenced organisms and those for which sequencing is in progress. All three main domains of life (bacteria, archaea, and eukaryota) are represented, as well as many viruses, phages, viroids, plasmids, and organelles.

Genome Reference Consortium (GRC)
The Genome Reference Consortium (GRC) maintains responsibility for the human and mouse reference genomes. Members consist of The Genome Center at Washington University, the Wellcome Trust Sanger Institute, the European Bioinformatics Institute (EBI) and the National Center for Biotechnology Information (NCBI). The GRC works to correct misrepresented loci and to close remaining assembly gaps. In addition, the GRC seeks to provide alternate assemblies for complex or structurally variant genomic loci. At the GRC website (http://www.genomereference.org), the public can view genomic regions currently under review, report genome-related problems and contact the GRC.

Glycans
A centralized page providing access and links to glycoinformatics and glycobiology related resources.

HIV-1, Human Protein Interaction Database

A database of known interactions of HIV-1 proteins with proteins from human hosts. It provides annotated bibliographies of published reports of protein interactions, with links to the corresponding PubMed records and sequence data.

Identical Protein Groups
A collection of consolidated records describing proteins identified in annotated coding regions in GenBank and RefSeq, as well as SwissProt and PDB protein sequences. This resource allows investigators to obtain more targeted search results and quickly identify a protein of interest.

Influenza Virus
A compilation of data from the NIAID Influenza Genome Sequencing Project and GenBank. It provides tools for flu sequence analysis, annotation and submission to GenBank. This resource also has links to other flu sequence resources, and publications and general information about flu viruses.

Journals in NCBI Databases
Subset of the NLM Catalog database providing information on journals that are referenced in NCBI database records, including PubMed abstracts. This subset can be searched using the journal title, MEDLINE or ISO abbreviation, ISSN, or the NLM Catalog ID.

MeSH Database
MeSH (Medical Subject Headings) is the U.S. National Library of Medicine's controlled vocabulary for indexing articles for MEDLINE/PubMed. MeSH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts.

MedGen
A portal to information about medical genetics. MedGen includes term lists from multiple sources and organizes them into concept groupings and hierarchies. Links are also provided to information related to those concepts in the NIH Genetic Testing Registry (GTR), ClinVar, Gene, OMIM, PubMed, and other sources.

NCBI C++ Toolkit Manual
A comprehensive manual on the NCBI C++ toolkit, including its design and development framework, a C++ library reference, software examples and demos, FAQs and release notes. The manual is searchable online and can be downloaded as a series of PDF documents.

NCBI Education Page
Provides links to tutorials and training materials, including PowerPoint slides and print handouts.

NCBI Glossary
Part of the NCBI Handbook, this glossary contains descriptions of NCBI tools and acronyms, bioinformatics terms and data representation formats.

NCBI Handbook
An extensive collection of articles about NCBI databases and software. Designed for a novice user, each article presents a general overview of the resource and its design, along with tips for searching and using available analysis tools. All articles can be searched online and downloaded in PDF format; the handbook can be accessed through the NCBI Bookshelf.

NCBI Help Manual
Accessed through the NCBI Bookshelf, the Help Manual contains documentation for many NCBI resources, including PubMed, PubMed Central, the Entrez system, Gene, SNP and LinkOut. All chapters can be downloaded in PDF format.

NCBI Pathogen Detection Project
A project involving the collection and analysis of bacterial pathogen genomic sequences originating from food, environmental and patient isolates. Currently, an automated pipeline clusters and identifies sequences supplied primarily by public health laboratories to assist in the investigation of foodborne disease outbreaks and discover potential sources of food contamination.

National Library of Medicine (NLM) Catalog
Bibliographic data for all the journals, books, audiovisuals, computer software, electronic resources and other materials that are in the library's holdings.

Nucleotide Database
A collection of nucleotide sequences from several sources, including GenBank, RefSeq, the Third Party Annotation (TPA) database, and PDB. Searching the Nucleotide Database will yield available results from each of its component databases.

Online Mendelian Inheritance in Man (OMIM)
A database of human genes and genetic disorders. NCBI maintains current content and continues to support its searching and integration with other NCBI databases. However, OMIM now has a new home at omim.org, and users are directed to this site for full record displays.

PopSet
Database of related DNA sequences that originate from comparative studies: phylogenetic, population, environmental and, to a lesser degree, mutational. Each record in the database is a set of DNA sequences. For example, a population set provides information on genetic variation within an organism, while a phylogenetic set may contain sequences, and their alignment, of a single gene obtained from several related organisms.

Protein Clusters
A collection of related protein sequences (clusters), consisting of Reference Sequence proteins encoded by complete prokaryotic and organelle plasmids and genomes. The database provides easy access to annotation information, publications, domains, structures, external links, and analysis tools.

Protein Database
A database that includes protein sequence records from a variety of sources, including GenPept, RefSeq, Swiss-Prot, PIR, PRF, and PDB.

Protein Family Models
A database that includes a collection of models representing homologous proteins with a common function. It includes conserved domain architecture, hidden Markov models and BlastRules. A subset of these models are used by the Prokaryotic Genome Annotation Pipeline (PGAP) to assign names and other attributes to predicted proteins.

PubChem BioAssay
Consists of deposited bioactivity data and descriptions of bioactivity assays used to screen the chemical substances contained in the PubChem Substance database, including descriptions of the conditions and the readouts (bioactivity levels) specific to the screening procedure.

PubChem Compound
Contains unique, validated chemical structures (small molecules) that can be searched using names, synonyms or keywords. The compound records may link to more than one PubChem Substance record if different depositors supplied the same structure. These Compound records reflect validated chemical depiction information provided to describe substances in PubChem Substance. Structures stored within PubChem Compounds are pre-clustered and cross-referenced by identity and similarity groups. Additionally, calculated properties and descriptors are available for searching and filtering of chemical structures.

PubChem Substance
PubChem Substance records contain substance information electronically submitted to PubChem by depositors. This includes any chemical structure information submitted, as well as chemical names, comments, and links to the depositor's web site.

PubMed
A database of citations and abstracts for biomedical literature from MEDLINE and additional life science journals. Links are provided when full text versions of the articles are available via PubMed Central (described below) or other websites.

PubMed Central (PMC)
A digital archive of full-text biomedical and life sciences journal literature, including clinical medicine and public health.

RefSeqGene
A collection of human gene-specific reference genomic sequences. RefSeq gene is a subset of NCBI's RefSeq database, and are defined based on review from curators of locus-specific databases and the genetic testing community. They form a stable foundation for reporting mutations, for establishing consistent intron and exon numbering conventions, and for defining the coordinates of other biologically significant variation. RefSeqGene is a part of the Locus Reference Genomic (LRG) Collaboration.

Reference Sequence (RefSeq)
A collection of curated, non-redundant genomic DNA, transcript (RNA), and protein sequences produced by NCBI. RefSeqs provide a stable reference for genome annotation, gene identification and characterization, mutation and polymorphism analysis, expression studies, and comparative analyses. The RefSeq collection is accessed through the Nucleotide and Protein databases.

Retrovirus Resources

A collection of resources specifically designed to support the research of retroviruses, including a genotyping tool that uses the BLAST algorithm to identify the genotype of a query sequence; an alignment tool for global alignment of multiple sequences; an HIV-1 automatic sequence annotation tool; and annotated maps of numerous retroviruses viewable in GenBank, FASTA, and graphic formats, with links to associated sequence records.

SARS CoV
A summary of data for the SARS coronavirus (CoV), including links to the most recent sequence data and publications, links to other SARS related resources, and a pre-computed alignment of genome sequences from various isolates.

Sequence Read Archive (SRA)
The Sequence Read Archive (SRA) stores sequencing data from the next generation of sequencing platforms including Roche 454 GS System®, Illumina Genome Analyzer®, Life Technologies AB SOLiD System®, Helicos Biosciences Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®.

Structure (Molecular Modeling Database)
Contains macromolecular 3D structures derived from the Protein Data Bank, as well as tools for their visualization and comparative analysis.

Taxonomy
Contains the names and phylogenetic lineages of more than 160,000 organisms that have molecular data in the NCBI databases. New taxa are added to the Taxonomy database as data are deposited for them.

Third Party Annotation (TPA) Database
A database that contains sequences built from the existing primary sequence data in GenBank. The sequences and corresponding annotations are experimentally supported and have been published in a peer-reviewed scientific journal. TPA records are retrieved through the Nucleotide Database.

Trace Archive
A repository of DNA sequence chromatograms (traces), base calls, and quality estimates for single-pass reads from various large-scale sequencing projects.

Viral Genomes
A wide range of resources, including a brief summary of the biology of viruses, links to viral genome sequences in Entrez Genome, and information about viral Reference Sequences, a collection of reference sequences for thousands of viral genomes.

Virus Variation
An extension of the Influenza Virus Resource to other organisms, providing an interface to download sequence sets of selected viruses, analysis tools, including virus-specific BLAST pages, and genome annotation pipelines.

Downloads

BLAST (Stand-alone)
BLAST executables for local use are provided for Solaris, LINUX, Windows, and MacOSX systems. See the README file in the ftp directory for more information. Pre-formatted databases for BLAST nucleotide, protein, and translated searches also are available for downloading under the db subdirectory.

FTP: BLAST Databases
Sequence databases for use with the stand-alone BLAST programs. The files in this directory are pre-formatted databases that are ready to use with BLAST.

FTP: CDD
This site provides full data records for CDD, along with individual Position Specific Scoring Matrices (PSSMs), mFASTA sequences and annotation data for each conserved domain. See the README file for full details.

FTP: ClinVar Data
This site provides full data extractions in XML and summary data in VCF format. It contains files with information about standard terms used in ClinVar, MedGen, and GTR.

FTP: FASTA BLAST Databases
Sequence databases in FASTA format for use with the stand-alone BLAST programs. These databases must be formatted using formatdb before they can be used with BLAST.

FTP: GenBank
This site contains files for all sequence records in GenBank in the default flat file format. The files are organized by GenBank division, and the full contents are described in the README.genbank file.

FTP: GenPept
The protein sequences corresponding to the translations of coding sequences (CDS) in GenBank are collected for each GenBank release..Please see the README file in the directory for more information.

FTP: Gene
This site contains three directories: DATA, GeneRIF and tools. The DATA directory contains files listing all data linked to GeneIDs along with subdirectories containing ASN.1 data for the Gene records. The GeneRIF (Gene References into Function) directory contains PubMed identifiers for articles describing the function of a single gene or interactions between products of two genes. Sample programs for manipulating gene data are provided in the tools directory. Please see the README file for details.

FTP: Gene Expression Omnibus (GEO) Profiles and Datasets
This site contains GEO data in two formats: SOFT (Simple Omnibus in Text Format) and MINiML (MIAME Notation in Markup Language). Summary text files and supplementary data are also available. Please see the README.TXT file for more information.

FTP: Genome
This site contains genome sequence and mapping data for organisms in Entrez Genome. The data are organized in directories for single species or groups of species. Mapping data are collected in the directory MapView and are organized by species. See the README file in the root directory and the README files in the species subdirectories for detailed information.

FTP: Genome Mapping Data
Contains directories for each genome that include available mapping data for current and previous builds of that genome.

FTP: NCBI Taxonomy
This site contains the full taxonomy database along with files associating nucleotide and protein sequence records with their taxonomy IDs. See the taxdump_readme.txt and gi_taxid.readme files for more information.

FTP: PubChem
This site provides data from the PubChem Substance, Compound and Bioassay databases for download via ftp. Full downloads of the databases are available along with daily, weekly and monthly updates for Substance and Compound. Substance and Compound data are provided in ASN.1, SDF and XML formats. See the README files for more information.

FTP: RefSeq
This site contains all nucleotide and protein sequence records in the Reference Sequence (RefSeq) collection. The ""release"" directory contains the most current release of the complete collection, while data for selected organisms (such as human, mouse and rat) are available in separate directories. Data are available in FASTA and flat file formats. See the README file for details.

FTP: SKY/M-Fish and CGH Data
This site contains SKY-CGH data in ASN.1, XML and EasySKYCGH formats. See the skycghreadme.txt file for more information.

FTP: SNP
Downloadable data for SNP.

FTP: Sequence Read Archive (SRA) Download Facility
This site contains next-generation sequencing data organized by the submitted sequencing project.

FTP: Site
FTP download site for NCBI databases, tools, and utilities.

FTP: Structure (MMDB)
This site contains ASN.1 data for all records in MMDB along with VAST alignment data and the non-redundant PDB (nr-PDB) data sets. See the README file for more information.

FTP: Trace Archive
This site contains the trace chromatogram data organized by species. Data include chromatogram, quality scores, FASTA sequences from automatic base calls, and other ancillary information in tab-delimited text as well as XML formats. See the README file for details.

FTP: UniVec
This site contains the UniVec and UniVec_Core databases in FASTA format. See the README.uv file for details.

FTP: Whole Genome Shotgun Sequences
This site contains whole genome shotgun sequence data organized by the 4-digit project code. Data include GenBank and GenPept flat files, quality scores and summary statistics. See the README.genbank.wgs file for more information.

FTP: dbGAP Open-Access Data
Open-access data generally include summaries of genotype/phenotype association studies, descriptions of the measured variables, and study documents, such as the protocol and questionnaires. Access to individual-level data, including phenotypic data tables and genotypes, requires varying levels of authorization.

MEDLINE (Leasing)
NLM leases MEDLINE/PubMed to U.S. individuals or organizations.

NCBI Data Specifications
Specifications for NCBI data in ASN.1 or DTD format are available on the Index of data_specs page. The "NCBI_data_conversion.html" links to the conversion tool.

National Library of Medicine (NLM) DTDs
A suite of tag sets for authoring and archiving journal articles as well as transferring journal articles from publishers to archives and between archives. There are four tag sets: Archiving and Interchange Tag Set - Created to enable an archive to capture as many of the structural and semantic components of existing printed and tagged journal material as conveniently as possible; Journal Publishing Tag Set - Optimized for archives that wish to regularize and control their content, not to accept the sequence and arrangement presented to them by any particular publisher; Article Authoring Tag Set - Designed for authoring new journal articles; NCBI Book Tag Set - Written specifically to describe volumes for the NCBI online libraries.

PubChem Download Service
This service allows users to download compound or substance records corresponding to a set of PubChem identifiers, which can be supplied manually or through a text file. Numerous download formats are available, including SDF, XML and SMILES.

PubMed Central (PMC) Open-Access Subset
The PMC Open-Access Subset is a relatively small part of the total collection of articles in PMC. Whereas the majority of articles in PMC are subject to traditional copyright restrictions, these articles are protected by copyright, but are made available under a Creative Commons or similar license that generally allows more liberal redistribution and reuse than a traditional copyright. Please refer to the license statement in each article for specific terms of use.

RSS Feeds
Subscribe to Web/RSS feeds for updates about NCBI resources.

Submissions

BioProject Submission
An online form that provides an interface for researchers, consortia and organizations to register their BioProjects. This serves as the starting point for the submission of genomic and genetic data for the study. The data does not need to be submitted at the time of BioProject registration.

ClinVar Submissions
Guidelines and instructions for submitting assertions about the pathogenicity of human genetic variants. These submissions can include summary data about a variant (variant level/aggregate data); support for variants per case (case-level) is in development.

Database of Genotype and Phenotype (dbGaP) Data Submission Policies
Guidelines and requirements for submitting genotype and phenotype association data to dbGaP.

GenBank: BankIt
A web-based sequence submission tool for one or a few submissions to the GenBank database, designed to make the submission process quick and easy.

GenBank: Barcode

Tool for submission to the GenBank database of Barcode short nucleotide sequences from a standard genetic locus for use in species identification.

GenBank: Sequin
A stand-alone software tool developed by the NCBI for submitting and updating entries to public sequence databases (GenBank, EMBL, or DDBJ). It is capable of handling simple submissions that contain a single short mRNA sequence, complex submissions containing long sequences, multiple annotations, segmented sets of DNA, as well as sequences from phylogenetic and population studies with alignments. For simple submission, use the online submission tool BankIt instead.

GenBank: tbl2asn
A command-line program that automates the creation of sequence records for submission to GenBank using many of the same functions as Sequin. It is used primarily for submission of complete genomes and large batches of sequences.

Gene Expression Omnibus (GEO) Web Deposit
Submit expression data, such as microarray, SAGE or mass spectrometry datasets to the NCBI Gene Expression Omnibus (GEO) database.

GeneRIF
GeneRIF provides a simple mechanism to allow scientists to add to the functional annotation of genes in the Gene database.

Genetic Testing Registry (GTR) Submissions
Guidelines and instructions for registering laboratories and submitting genetic test information including clinical and research tests for germline or somatic test targets. GTR welcomes registration of cytogenetic, biochemical, and molecular tests for Mendelian disorders, pharmacogenetic phenotypes and complex panels.

NIH Manuscript Submissions (NIHMS)
The NIH Manuscript Submission (NIHMS) System is used to submit manuscripts that arise from NIH funding to the PubMed Central digital archive, in accordance with the NIH Public Access Policy and the law it implements. The law and Public Access Policy are intended to ensure that the public has access to the published results of NIH-funded research.

PubChem Upload
This site enables users to submit data to the PubChem Substance and BioAssay databases, including chemical structures, experimental biological activity results, annotations, siRNA data and more. It can also be used to update previously submitted records.

SNP Submission Tool
The SNP database tools page provides links to the general submission guidelines and to the submission handle request. The page has also two specific links for single- or batch submissions of the human variation data using Human Genome Variation Society nomenclature.

Sequence Read Archive Submission
This link describes how submitters of SRA data can obtain a secure NCBI FTP site for their data, and also describes the allowed data formats and directory structures.

Submission Portal
A single entry point for submitters to link to and find information about all of the data submission processes at NCBI. Currently, this serves as an interface for the registration of BioProjects and BioSamples and submission of data for WGS and GTR. Future additions to this site are planned.

Trace Archive Submission
This link describes how submitters of trace data can obtain a secure NCBI FTP site for their data, and also describes the allowed data formats and directory structures.

Tools

Amino Acid Explorer
This tool allows users to explore the characteristics of amino acids by comparing their structural and chemical properties, predicting protein sequence changes caused by mutations, viewing common substitutions, and browsing the functions of given residues in conserved domains.

BLAST Microbial Genomes

Performs a BLAST search for similar sequences from selected complete eukaryotic and prokaryotic genomes.

BLAST RefSeqGene
Performs a BLAST search of the genomic sequences in the RefSeqGene/LRG set. The default display provides ready navigation to review alignments in the Graphics display.

BLAST Tutorials and Guides
This page links to a number of BLAST-related tutorials and guides, including a selection guide for BLAST algorithms, descriptions of BLAST output formats, explanations of the parameters for stand-alone BLAST, directions for setting up stand-alone BLAST on local machines and using the BLAST URL API.

Basic Local Alignment Search Tool (BLAST)
Finds regions of local similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as to help identify members of gene families.

Batch Entrez
Allows you to retrieve records from many Entrez databases by uploading a file of GI or accession numbers from the Nucleotide or Protein databases, or a file of unique identifiers from other Entrez databases. Search results can be saved in various formats directly to a local file on your computer.

CDTree
A stand-alone application for classifying protein sequences and investigating their evolutionary relationships. CDTree can import, analyze and update existing Conserved Domain (CDD) records and hierarchies, and also allows users to create their own. CDTree is tightly integrated with Entrez CDD and Cn3D, and allows users to create and update protein domain alignments.

COBALT
COBALT is a protein multiple sequence alignment tool that finds a collection of pairwise constraints derived from conserved domain database, protein motif database, and sequence similarity, using RPS-BLAST, BLASTP, and PHI-BLAST.

Cn3D
A stand-alone application for viewing 3-dimensional structures from NCBI's Entrez retrieval service. Cn3D runs on Windows, Macintosh, and UNIX and can be configured to receive data from most popular web browsers. Cn3D simultaneously displays structure, sequence, and alignment, and has powerful annotation and alignment editing features.

Coffee Break
Part of the NCBI Bookshelf, Coffee Break combines reports on recent biomedical discoveries with use of NCBI tools. Each report incorporates interactive tutorials that show how NCBI bioinformatics tools are used as a part of the research process.

Conserved Domain Architecture Retrieval Tool (CDART)
Displays the functional domains that make up a given protein sequence. It lists proteins with similar domain architectures and can retrieve proteins that contain particular combinations of domains.

Conserved Domain Search Service (CD Search)
Identifies the conserved domains present in a protein sequence. CD-Search uses RPS-BLAST (Reverse Position-Specific BLAST) to compare a query sequence against position-specific score matrices that have been prepared from conserved domain alignments present in the Conserved Domain Database (CDD).

E-Utilities
Tools that provide access to data within NCBI's Entrez system outside of the regular web query interface. They provide a method of automating Entrez tasks within software applications. Each utility performs a specialized retrieval task, and can be used simply by writing a specially formatted URL.

Ebot
A tool that allows users to construct an E-utility analysis pipeline using an online form, and then generates a Perl script to execute the pipeline.

Gene Expression Omnibus (GEO) BLAST
Tool for aligning a query sequence (nucleotide or protein) to GenBank sequences included on microarray or SAGE platforms in the GEO database.

Genetic Codes

Displays the genetic codes for organisms in the Taxonomy database in tables and on a taxonomic tree.

Genome BLAST
This tool compares nucleotide or protein sequences to genomic sequence databases and calculates the statistical significance of matches using the Basic Local Alignment Search Tool (BLAST) algorithm.

Genome Data Viewer (GDV)
A genome browser for interactive navigation of eukaryotic RefSeq genome assemblies with comprehensive inspection of gene, expression, variation and other annotations. GDV offers easy-to-load analytical track pre-configurations, a menu of data tracks for easy display and customization, and supports upload and analysis of user data. This browser also enables the production of displays for publishing.

Genome Decoration Page
An online tool that assists in the production of journal quality figures of annotations on an ideogram or sequence representation of an assembly.

Genome Remapping Service
NCBI's Remap tool allows users to project annotation data and convert locations of features from one genomic assembly to another or to RefSeqGene sequences through a base by base analysis. Options are provided to adjust the stringency of remapping, and summary results are displayed on the web page. Full results can be downloaded for viewing in NCBI's Genome Workbench graphical viewer, and annotation data for the remapped features, as well as summary data, is also available for download.

Genome Workbench
An integrated application for viewing and analyzing sequence data. With Genome Workbench, you can view data in publically available sequence databases at NCBI, and mix these data with your own data.

LinkOut
A service that allows third parties to link directly from PubMed and other Entrez database records to relevant web-accessible resources beyond the Entrez system. Examples of LinkOut resources include full-text publications, biological databases, consumer health information and research tools.

Map Viewer
Provides special browsing capabilities of maps and assembled sequences for a subset of organisms. You can view and search an organism's complete genome, display maps, and zoom into progressively greater levels of detail, down to the sequence data for a region of interest.

Multiple Sequence Alignment Viewer
An interactive web application that enables users to visualize multiple alignments created by database search results or other software applications. The MSA Viewer allows users to upload an alignment and set a master sequence, and to explore the data using features such as zooming and changing of coloration.

NCBI News
Provides information on new and updated resources and NCBI research and development projects. The News site contains feature articles highlighting services, resource features and tools, as well as frequent postings describing important announcements regarding key datasets and services of interest to the user community. Links to NCBI's social media sites along and a list of available RSS feeds and Email listservs are provided.

NCBI Toolbox
A set of software and data exchange specifications used by NCBI to produce portable, modular software for molecular biology. The software in the Toolbox is primarily designed to read records in Abstract Syntax Notation 1 (ASN.1) format, an International Standards Organization (ISO) data representation format.

OSIRIS
A public domain quality assurance software package that facilitates the assessment of multiplex short tandem repeat (STR) DNA profiles based on laboratory-specific protocols. OSIRIS evaluates the raw electrophoresis data using an independently derived mathematically-based sizing algorithm. It offers two new peak quality measures - fit level and sizing residual. It can be customized to accommodate laboratory-specific signatures such as background noise settings, customized naming conventions and additional internal laboratory controls.

Open Reading Frame Finder (ORF Finder)
A graphical analysis tool that finds all open reading frames in a user's sequence or in a sequence already in the database. Sixteen different genetic codes can be used. The deduced amino acid sequence can be saved in various formats and searched against protein databases using BLAST.

PSSM Viewer

Allows users to display, sort, subset and download position-specific score matrices (PSSMs) either from CDD records or from Position Specific Iterated (PSI)-BLAST protein searches. The tool also can align a query protein to the PSSM and highlight positions of high conservation.

Phenotype-Genotype Integrator (PheGenI)
Supports finding human phenotype/genotype relationships with queries by phenotype, chromosome location, gene, and SNP identifiers. Currently includes information from dbGaP, the NHGRI GWAS Catalog, and GTeX. Displays results on the genome, on sequence, or in tables for download.

Primer-BLAST
The Primer-BLAST tool uses Primer3 to design PCR primers to a sequence template. The potential products are then automatically analyzed with a BLAST search against user specified databases, to check the specificity to the target intended.

ProSplign
A utility for computing alignment of proteins to genomic nucleotide sequence. It is based on a variation of the Needleman Wunsch global alignment algorithm and specifically accounts for introns and splice signals. Due to this algorithm, ProSplign is accurate in determining splice sites and tolerant to sequencing errors.

PubChem Power User Gateway (PUG)
PUG provides access to PubChem services via a programmatic interface. PUG allows users to download data, initiate chemical structure searches, standardize chemical structures and interact with the E-utilities. PUG can be accessed using either standard URLs or via SOAP.

PubChem Standardization Service
Standardization, in PubChem terminology, is the processing of chemical structures in the same way used to create PubChem Compound records from contributors' original structures. This service lets users see how PubChem would handle any structure they would like to submit.

PubChem Structure Search
PubChem Structure Search allows the PubChem Compound Database to be queried by chemical structure or chemical structure pattern. The PubChem Sketcher allows a query to be drawn manually. Users may also specify the structural query input by PubChem Compound Identifier (CID), SMILES, SMARTS, InChI, Molecular Formula, or by upload of a supported structure file format.

PubMed Clinical Queries
A specialized PubMed search form targeted to clinicians and health services researchers. The page simplifies searching by clinical study category, finding systematic reviews and searching the medical genetics literature.

PubMed Tutorials
A collection of web and flash tutorials on PubMed searching and linking, saving searches in MyNCBI, using MeSH and other PubMed services.

Related Structures
The Related Structures tool allows users to find 3D structures from the Molecular Modeling Database (MMDB) that are similar in sequence to a query protein. Although the query protein may not yet have a resolved structure, the 3D shape of a similar protein sequence can shed light on the putative shape and biological function of the query protein.

SNP Database Specialized Search Tools
A variety of tools are available for searching the SNP database, allowing search by genotype, method, population, submitter, markers and sequence similarity using BLAST. These are linked under ""Search"" on the left side bar of the dbSNP main page.

Sequence Viewer
Provides a configurable graphical display of a nucleotide or protein sequence and features that have been annotated on that sequence. In addition to use on NCBI sequence database pages, this viewer is available as an embeddable webpage component. Detailed documentation including an API Reference guide is available for developers wishing to embed the viewer in their own pages.

Splign
A utility for computing cDNA-to-Genomic sequence alignments. It is based on a variation of the Needleman-Wunsch global alignment algorithm and specifically accounts for introns and splice signals. Due to this algorithm, Splign is accurate in determining splice sites and tolerant to sequencing errors.

Taxonomy Browser
Supports searching the taxonomy tree using partial taxonomic names, common names, wild cards and phonetically similar names. For each taxonomic node, the tool provides links to all data in Entrez for that

node, displays the lineage, and provides links to external sites related to the node.

Taxonomy Common Tree
Generates a taxonomic tree for a selected group of organisms. Users can upload a file of taxonomy IDs or names, or they can enter names or IDs directly.

Taxonomy Statistics
Displays the number of taxonomic nodes in the database for a given rank and date of inclusion.

Taxonomy Status Reports
Displays the current status of a set of taxonomic nodes or IDs.

Tree Viewer
A tool for creating and displaying phylogenetic tree data. Tree Viewer enables analysis of your own sequence data, produces printable vector images as PDFs, and can be embedded in a webpage.

Variation Viewer
A genomic browser to search and view genomic variations listed in dbSNP, dbVar, and ClinVar databases. Searches can be performed using chromosomal location, gene symbol, phenotype, or variant IDs from dbSNP and dbVar. The browser enables exploration of results in a dynamic graphical sequence viewer with annotated tables of variations.

VecScreen
A system for quickly identifying segments of a nucleic acid sequence that may be of vector origin. VecScreen searches a query sequence for segments that match any sequence in a specialized non-redundant vector database (UniVec).

Vector Alignment Search Tool (VAST)
A computer algorithm that identifies similar protein 3-dimensional structures. Structure neighbors for every structure in MMDB are pre-computed and accessible via links on the MMDB Structure Summary pages. These neighbors can be used to identify distant homologs that cannot be recognized by sequence comparison alone.

Viral Genotyping Tool
This tool helps identify the genotype of a viral sequence. A window is slid along the query sequence and each window is compared by BLAST to each of the reference sequences for a particular virus.

How To

- Save text searches and set up automated searches with E-mailed results
- Submit data to NCBI
- Find information on glycans and glycobiology resources
- Find bioassays in which a given drug is active
- Find bioassays that test a particular disease or protein target
- Submit sequence data to NCBI
- Download NCBI Software
- Retrieve all sequences for an organism or taxon
- Find the function of a gene or gene product
- Find expression patterns
- Find genes associated with a phenotype or disease
- Find human variations associated with a phenotype or disease (clinical association)
- Find sequenced genomes, including those in progress, for a taxonomic group
- Compare protein homologs between two microbial genomes
- Download the complete genome for an organism
- Convert feature coordinates between genomic assemblies
- Find a homolog for a gene in another organism
- Obtain the full text of an article
- Find articles about a topic similar to that in a given article
- Find transcript sequences for a gene
- View the 3D structure of a protein
- Find a curated version of a sequence record (NCBI Reference Sequence)
- Align two or more 3D structures to a given structure
- Find published information on a gene or sequence
- Run BLAST software on a local computer
- Submit multiple query sequences in a single BLAST search
- Automate BLAST searches performed on NCBI servers
- Design PCR primers and check them for specificity

- Compare your sequence to the RefSeqGene/LRG standard
- Generate a Common Tree for a set of taxa
- Find the complete taxonomic lineage for an organism
- Learn about an NCBI resource
- Learn about the basics of molecular biology and bioinformatics
- Find out what's new at NCBI
- View a mutation site in a 3D structure
- View genotype frequency data for a gene, disease or short genetic variation
- Download a large, custom set of records from NCBI